TASTE PANELS

by

George Emery Ferris

March 1, 1956

BU-61-M

CONTENTS

I. CLASSIFICATION AND PURPOSES OF TASTE PANELS

Organoleptic panels, in general, have to be used when the quality of the food product being investigated cannot be fully evaluated in terms of objective or physical measurements. Hence such qualities as flavor and odor acceptability, mouthfeel, and "body" have almost invariably to be judged subjectively, and it is often convenient -- though not necessary -- to evaluate other characteristics, e.g., texture, color intensity and acceptability, viscosity, clarity, consistency, dry appearance, fattiness, and other special properties of the particular product, the same way.

Until a few years ago the tendency was to replace subjective methods by objective ones as far as possible, because:
(a) the distinction between consumer and analytic panels was not understood and confused;
(b) conditions in store-testing and home-use-testing on the one side, and in laboratory testing on the other side, were not standardized;
(c) the method of evaluating results and design of experiments had fallen into the hands of methodologists, whose knowledge was insufficient and brought the methods into disrepute. More recently, however, it has been realized that panels are capable of accuracy under optimum conditions quite comparable to physical tests (see Dove (17), who asserts they can have accuracy of a "microbalance"), that there are occasions when their use is inevitable; conditions for consumer and analytic panels have been standardized (see e.g. Deming (16) and Hopkins (28), respectively), and food companies now employ consultants well-versed in the application of statistics to food problems.

For the purposes of the discussions below it will be best to treat consumer panels and analytic panels separately. The manufacturer or research worker will at various stages of development of a food product be interested in running both types of test. In the past he has not formulated the question he wanted answered before deciding which type of panel to avail himself of, nor did he realize that to answer specific questions corresponding specific designs were at his disposal both within the framework of consumer panels and analytic panels. The former should be used when it is necessary to obtain information on the acceptability of the product either in an absolute way, or

in relation to any competitors that may be on the market already. The desire, then, is to establish absolute or relative preference.

Analytic panels will be used to establish differences. These differences may be very minor and quite unimportant from the point of view of acceptability to the consumer. But such panels are of tremendous use to the manufacturer in conducting quality control panels, experiments on the best method of storage, optimum composition experiments, etc. As always in classification, there will be some situations intermediate to these, and though not recommended, mention is made of "dual-purpose" panels in the last chapter.

II. CONSUMER PANELS

(a) Purpose

The purpose of a consumer test is to establish the absolute or relative acceptability of a given product with regard to one or more (usually all the relevant) characteristics. The optimum for any one characteristic may be at an extremity or at the midpoint of a given scale (e.g., uniformity in texture of bread versus fattiness in meat). From the economic point of view the problem may be best illustrated in the simple case of two competing products, A and B, say. We wish to determine, with reasonable pre-specified accuracy, the percentage of the population of consumers of these products who

(i) either cannot differentiate between A and B, or if they can, have no preference; let this percentage be π_o;

(ii) can differentiate and prefer A; let this percentage be π_A;

(iii) can differentiate and prefer B; let this percentage be π_B.

The situation corresponds to determining a point on the triangle below, with a confidence region around it, so as to decide on the course of action to be followed by the manufacturer. If the estimated position falls in region I, the cheaper of A and B is marketed. If it falls in II, A only, if in IV, B only is marketed. If the point falls in III, both products need marketing and suitable advertising.

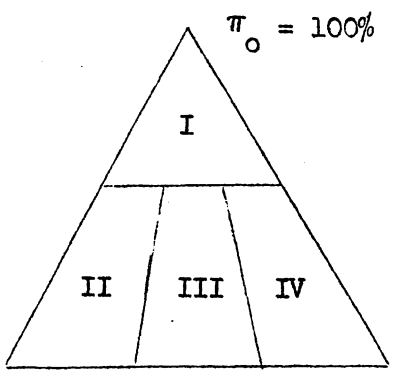

Diagram I

If the point and its confidence region overlap the borders of these regions, the course involving the smallest possible loss should be followed.

Consumer tests should be so designed and carried out as to answer this formulation (or a similar formulation if more than two products are involved, as in general), and to do so in the most efficient and convenient way.

(b) Method of Selection

A consumer panel must be selected bearing in mind the following considerations:

(1) It must be large, heterogeneous, representative of the population consuming that product which is being investigated (obviously not the same for champagne as for bread, nor the same for chili-con-carne as corn bread).

(2) Its members must not be trained or over-instructed.

(3) An estimation of the number of guessers must be possible, because repeated evidence is available that panel-members tend to vote for preferences "just to oblige",

To add to some of these: Obviously no attempt should be made to exclude non-discriminators on grounds either genetic (taste-blind), medical, nor any similar reason, if in the first place these members were selected by a random or other statistical process (see (c) below), as (1) would be violated otherwise. In view of (2) the use of "standing home-use-panels" is much to be discouraged, as members may get over-familiar with the product; i.e., renew the panel at the end of each series of tests. Finally, (3) makes necessary a "visit-revisit" technique or the "double-pair" method described in (c) below.

(c) Statistical Methods

The principal difficulty is in defining the population and obtaining a representative sample therefrom. Deming (16) and other books on Sample Survey show how a representative panel may be chosen. The multitude of biases which may be introduced by mailed questionnaires, interviewers, and similar techniques are also pointed out there. As to the actual statistical design to be used, one may consult certain "standard" techniques, e.g., "ranking tests" (see Boggs and Hanson (5)), "paired tests" (see ibid, also Cover (10), (11)), "triangle", "triad", "delta", "triple comparison" tests (see Boggs and Hanson (5), Hening (26), or Roessler et al. (36)). The latter may be combined with preference as in Helm and Trolle (25). One may use "difference-preference" tests

(see Metzner (32)), the "constant stimulus" method (see Metzner (32)) or "matching with standards" (see Handschumaker (24), or Punnet and Eddy (35)). Occasionally more convenient methods like the show of hands at dinner (see Blakeslee (4)), or the judgment by amounts left on plates (see Tomkins (40)), or the setting up of vending machines selling competing products by the "select your preference" technique, may suggest themselves. The only method known to the author, where the parameters π_o, π_A, π_B, of Chapter II(a) above can be estimated unbiassed is the "double-pair" method (see Ferris (21)) or a less efficient version, known as the "duo-trio" method (see Bradley (6)). Another type of method statistically is the method of "paired comparisons" (see Scheffe (37) or Mosteller (33)), but these are very subject to biases in the form of the questionaire.

(d) Store-Test versus Home-Use-Test

The store-test method of consumer testing is preferred by most manufacturers because of its convenience and rapidity, but thought should be given whether such tests meet the stipulations of Chapter II(c) above as regards representativeness. The poorest and richest consumers hardly do their own shopping. Also, where subtle flavors or odors are involved the atmosphere of supermarkets may not be conducive to best results. In stores the "double-pair" method is recommended. The home-use-tests have several advantages, viz., that the product will be consumed at the usual meal-time, prepared the usual way, and in an atmosphere conducive to good evaluation. The home-use-panel may be tested by the "visit-revisit" method. A representative sample is, at least theoretically, easily obtained. Families may, when applicable, be used as sampling units. Care has to be exercised in not using the same panel for too long (they become "trained"), and to allow in the statistical evaluation for a change in preference during a series (especially important in the case of new or highly flavored products).

III. ANALYTIC PANELS

(a) Purpose and Functions

The purpose of analytic panels, usually set up in the laboratory of the manufacturer, is to detect minor differences in the flavor (or other quality) of the food product. The differences detected may be so small as to be unimportant from the point of view of acceptability to the consumer. Two very common purposes are given in illustration:

(1) Quality Control Panel: In such cases a small panel of 6 or 10 (or even less) members will meet daily to assess a representative sample of that day's production. They may judge any quality or qualities not readily assessable by objective means. The panel as a body functions as a piece of equipment. Its results may not be reproducible elsewhere, but this does not matter. They will stay consistent with themselves. Naturally, the constitution of the panel should be constant, and members should be highly trained.

(2) Storage Studies: Suppose different methods of storage are under study. After the material has been randomly allotted to the various storage-treatments, at various intervals random samples will be withdrawn from storage and judged for several characteristics by the analytic panel. Minor differences, if significant, may show the superiority of one method of storage over another, without implying that even the worst method of storage is unacceptable to consumers. If long periods are involved in storage, the composition of the panel may change -- as will the mental standards of the unchanged judges -- so that it is usually more advisable to analyze regression coefficients of scores for various qualities on time, than to analyze the scores themselves either at any one withdrawal or pooling data over withdrawals, because by the regression method uncontrolled factors are pushed into the error and not allowed to bias conclusions.

Naturally, analytic panels serve many other purposes, e.g., the study of the effect of small variations in the composition of food products on organoleptic characteristics, or the effect of varying the source of ingredients, the method of combining the ingredients, or the effect of the time or temperature at which the product is served.

(b) Panel Selection

The following factors need to be considered in selecting an analytic panel:

(1) Experience
(2) Availability
(3) Age
(4) Sex
(5) Health

It has been demonstrated repeatedly that experienced judges are more efficient in every sense of the word in taste-testing of a great variety of products (see Helm and Trolle (25)). Because of the importance of keeping the composition of the panel constant, availability is of primary importance (see Briant (7), Greenwood et al. (23), Lowe (31), Tomkins (40)). Several experiments report on age and sex differences in judges as regards thresholds for the various primary tastes. Some even investigate professional differences. Many of these investigators try to generalize their conclusions unduly, and it has yet to be demonstrated that judges with low thresholds are more adept than others at detecting differences at higher levels of concentration and in mixtures of primary tastes (see Metzner (32), Hopkins (28), for age differences, and Anonymous (1) for sex differences). Certain diseases and habits should disqualify persons from acting as panel members and judges should be physically well, not fatigued or worried (see Clendenning (9)). Many psychological factors should be considered and controlled, even the type of container for holding the samples, and the interest and cooperation of panel members must be gained and maintained (see Crist and Seaton (13)). Care must be taken to select the panel by tests on the food in question, to ensure inbred likes and dislikes are detected in time. Certain genetic disabilities can also be discovered (see Blakeslee (3)). The reliability of judges can be checked by several methods:

(1) Recognition of duplicates
(2) Ranking increasing concentrations correctly
(3) Scores on duplicate samples
(4) Deviation from panel average
(5) Use of end or midpoint reference standards
(6) Questionnaires to detect eccentricities
(7) Control chart methods

These are reported by Dove (18), Peret (34) and others.

The size of panel required should be determined by the statistician according to the degree of accuracy required. Panels of sizes 3 to 50 are reported (see Anonymous (1), Crist and Seaton (13), Hicks (27), Jacobs (29)). It will naturally vary with the character of the product and the type of study.

(c) Training

Training procedures using solutions of primary tastes, flavor essences and reference samples of predetermined scores are reported (see Weckel (41)). Some long period procedures and procedures involving recognition of off-flavors in particular products are also available (see Sharp (38)).

(d) Checking Performance

In Chapter III(b) above were listed some techniques for evaluating the reliability of judges. In addition the statistical techniques involving analyses of deviations on replicate samples and of individual scores, examination of the individual judge's regression on, or correlations with, the panel mean, and finally the technique of multiple choice tests are all useful "tricks of the trade" (see Hening (26), Ziegler and Schofield (42)).

(e) Preparation of Samples

The size, temperature, method of cooking or other preparation of samples, also the method of serving need of course to be treated separately for each type of product. In Dawson (15) pp. 19 et seq. a guide is given in respect to these questions product by product, and references are available to previous taste-tests with the product in question. Where possible, conclusions are generalized.

(f) Conditions of Judging and Judging Room

The following factors are of importance:

(1) Time of day
(2) Utensils used
(3) Coding of samples
(4) Time after smoking
(5) Discussions allowed at judging session
(6) Time allowed for tasting
(7) Method of removing flavors from mouth
(8) Location, ventilation, lighting, temperature control of judging room.
(9) Seating arrangement.

The two most useful texts for clarifying these mechanical details and attempting to standardize conditions in the laboratory for taste-testing are Hopkins (28) and Dawson (15) pp. 27 et seq., and from the latter many further references may be obtained.

(g) Factors Determining Accuracy of Tests

(1) The number and kind of characteristics evaluated
(2) Uniformity and quality of the material
(3) Standardization of terminology to describe quality
(4) Number of samples, replications
(5) Use of reference-standards
(6) Amount of information given the panel
(7) Scheduling of samples for concurrent testing

Decisions on these questions should be made in conjunction with the statistician. At the same time decision has to be made on such further points as:

(8) Should descriptive adjectives be used?
(9) Should numerical scores only be used?
(10) Should ranking be used?
(11) Should paired tests be used?
(12) Should triangle, triad, delta, triple comparison, duo-trio, or multiple choice tests be used?
(13) Should dilution tests be used?
(14) Should difference-preference tests be used?
(15) Should matching or some other tests be used?
(16) What type of score card should be used?
(17) How many characteristics should be used?
(18) How many points on the scale for each characteristic should be used?
(19) Should scores be combined? If so, what weight should each characteristic have? Are the scores additive?
(20) Is it better to ask judges to score by a graphical method (mark a spot on a scale) rather than numerically?
(21) Do the scores or ranks need transformation before analysis?
(22) Will each judge's scores be analysed separately, or in lump for the panel?

Information on most of these points is available in Dawson (15) pp. 32 et seq., and more recently in Bradley (6). No generalization can be attempted, each product and study having to be subjected to separate examination.

(h) Correlation with Objective Tests

Where the correlation between some easily measured physical or chemical property of the particular product with one of the qualities usually evaluated by organoleptic means has been shown to be high, in a preliminary experiment, say, the objective measurement may be used in place of the subjective one in later work. Such correlations have been reported in many instances. Protein content and flavor of bread (see Davis and Halliday (14)), water absorption and preference for cakes (Swartz (39)), pH of wafers and retention of flavor (coumarin)(see Gilmer et al. (22)) only to mention cereal products. Reports of similar relation in Beverages, Dairy Products, Eggs, Fats and Oils, Fruits, Meats, Poultry, Vegetables, Primary Tastes and Miscellaneous Foods are available in Dawson (15) pp. 35 et seq., where explanations are also given of the significance of the correlation in these cases.

(i) Statistical Design

All the usual statistical designs, viz. completely randomized design, randomized complete blocks, latin squares, factorial experiments, partially and completely confounded designs, fractional replication, split block and split plot designs, and particularly incomplete block designs, are of use in taste-testing. These designs will be found described in Federer (20), and in application in Cox (12), whilst some useful models specific to taste-testing may be found in Ferris (21) and Calvin (8). Useful statistical summaries are Boggs (5) and the Bi-Annual Reports of the Virginia Agricultural Experiment Station on Statistical Methods for Sensory Difference Tests of Food Quality (6). For some reason "multiple choice" designs waned in popularity until the advent of the triangle design (see Bengtsson and Helm (2)); since then triangle designs have been used to excess, where other multiple choice designs would be more efficient of time and material.

(j) Methods of Analysis

Appropriate to the designs discussed in (i) above there are the usual statistical methods of analysis. It may be convenient to analyse averages, ranges, percentages, ratios, regression or correlation coefficients; the usual techniques of Chi-square, t-tests, analyses of variance, control chart methods or discriminant functions may apply directly or with certain modifications. There is quite some quantity of literature on non-parametric methods, mostly

ranking (see Durbin (19)) and there is a very great deal of work to be done in adapting statistical models applicable to other fields of application to application in taste-testing (see Calvin (8) Ferris (21)).

IV. "DUAL-PURPOSE" PANELS

Attempts are often made in the laboratories of various food concerns to set up so-called "dual-purpose" panels. The intention is to use a panel of employees to judge differences and indicate preferences simultaneously. For the purposes of statistical analysis the fictitious assumption is made "in order to obtain a first approximation" that the panel so chosen is representative of the consumer population. The analysis is performed in two stages -- first the total scores of the panel are analysed, just as if it were functioning as an analytic panel; at a second stage by the "split-plot" method the individual scores of judges and the interactions are analysed. A good example of this procedure is given in Kempthorne (30) and in Ferris (21). Whereas the procedure is not recommended, there are times when its use is fairly inevitable.

BIBLIOGRAPHY

(1) Anonymous. How Do We Maintain Sponsored Brand Quality? Safeway News 3 (8 and 9), 12-15, 1948.

(2) Bengtsson, K. and Helm, E. Principles of Taste Testing. Wallerstein Labs. Commun. 9, 171-180, 1946.

(3) Blakeslee, A. F. Genetics of Sensory Thresholds: Taste for Phenyl-Thio-Carbamide. Natl. Acad. Sci. Proc. 18, 120-130, 1932.

(4) Blakeslee, A. F. A Dinner Demonstration of Threshold Differences in Taste and Smell. Science (n.s.) 81, 504-507, 1935.

(5) Boggs, M. M. and Hanson, H. L. Analysis of Foods by Sensory Difference Tests. Advances in Food Research, 2, 219-258, 1949.

(6) Bradley, R. A. (Editor) Statistical Methods for Sensory Difference Tests of Food Quality. Bi-Annual Reports of Virginia Agricultural Experiment Station, e.g. No. 7, Dec. 1953.

(7) Briant, A. M. Methods for Judging Flavor and Odor of Cooked Potatoes from Soils Treated with Wire Worm Control. Amer. Potato Jour. 26, 300-304, 1949.

(8) Calvin, L. D. Doubly Balanced Incomplete Block Designs for Experiments in Which the Treatment Effects are Correlated. Biometrics 10, (1) 61-88, 1954.

(9) Clendenning, T. Flavor in Confections. Mfg. Confectioner 20 (2), 23-25, 1940.

(10) Cover, S. A New Subjective Method of Tasting Tenderness in Meat -- the Paired-Eating Method. Food Res. 1, 287-295, 1936.

(11) Cover, S. Some Modifications of the Paired-Eating Method in Meat Cookery Research. Food Res. 5, 379-394, 1940.

(12) Cox, G. M. Statistics as a Tool for Research. Jour. Home Econ. 36, 575-580, 1944.

(13) Crist, J. W. and Seaton, H. L. Reliability of Organoleptic Tests. Food Res. 6, 529-536, 1941.

(14) Davis, M. V. and Halliday, E. G. Studies on All-Purpose Flour. Cereal Chem. 16, 414-418, 1939.

(15) Dawson, E. H. Sensory Methods for Measuring Differences in Food Quality. Ag. Inf. Bull. 34, U.S.D.A., 1951.

(16) Deming, W. E. Some Theory of Sampling. John Wiley and Sons Inc., New York, 1950.

(17) Dove, W. F. Developing Food Acceptance Research. Science 103, 187-190, 1946.

(18) Dove, W. F. Food Acceptability -- Its Determination and Evaluation. Food Technol. 1, 39-50, 1947.

(19) Durbin, V. Incomplete Blocks in Ranking Experiments. Brit. Jour. Psych. (Stat. Sect.) 4, 85-90, 1943.

(20) Federer, W. T. Experimental Design. Theory and Application. Macmillan Co., New York, 1956.

(21) Ferris, G. E. Statistical Designs in Taste-Testing. Cornell University, Ph.D. Thesis, 1956.

(22) Gilmer, R. S., Kinder, D. E. and Bohn, R. M. The Disappearance of Flavors in Biscuits. Cereal Chem. 13, 421-427, 1936.

(23) Greenwood, M. L. and Tice, J. M. Palatability Tests on Potatoes Grown in Soil Treated with Insecticides Benzene Hexachloride, Chlordane, and Chlorinated Camphene. Jour. Agr. Res. 78, 477-482, 1949.

(24) Handschumaker, E. A Technique for Testing the Reversion Properties of Hydrogenated Soybean Oil Shortenings. Amer. Oil Chem. Soc. Jour. 25, 54-56, 1948.

(25) Helm, E. and Trolle, B. Selection of a Taste Panel. Wallerstein Lab. Commun. 9, 181-194, 1946.

(26) Hening, J. C. Flavor Evaluation Procedures. N. Y. State Agr. Expt. Sta. Tech. Bul. 284, 1-20, 1948.

(27) Hicks, E. W. Tasting Tests. Food Preserv. Quart. 8(1), 1-5, 1948.

(28) Hopkins, J. W. Precision of Assessment of Palatability of Foodstuffs by Laboratory Panels. Canad. Jour. Res. Sect. F, Technol. 24, 203-214, 1946.

(29) Jacobs, M. B. Flavor Measurement. Amer. Perfumer and Essential Oil Rev. 52, 133, 135, 1948.

(30) Kempthorne, O. The Design and Analysis of Experiments. Wiley and Sons Inc., New York, 1952.

(31) Lowe, B. and Stewart, G. F. Subjective and Objective Tests as Food Research Tools. Food Technol. 1, 30-38, 1947.

(32) Metzner, C. A. Investigation of Odor and Taste. Psychological Principles. Wallerstein Labs. Commun. 6, 5-18, 1943.

(33) Mosteller, F. Remarks on the Method of Paired Comparisons. Psychometrika 16, 3-10, 1951.

(34) Peret, H. Taste Test Panels. Nat'l. Provisioner 121, (20), 12-13, 1949.

(35) Punnett, P. W. and Eddy, W. F. What Flavor Measurement Reveals About Keeping Coffee Fresh. Food Indus. 2, 401-404, 1930.

(36) Roessler, E. B., Warren, J. and Guymon, J. F. Significance in Triangular Taste Tests. Food Res. 13, 503-505, 1948.

(37) Scheffe', H. A. An Analysis of Variance for Paired Comparisons. Jour. Amer. Stat. Assoc. 47, No. 259, 381-400, 1951.

(38) Sharp, P. F. Some of the Questions Raised in an Address on Factors Influencing the Flavor of Milk. Milk Plant Monthly 30 (2), 31-34, 1941.

(39) Swartz, V. W. Two Further Simple Objective Tests for Judging Cake Quality. Cereal Chem. 15, 247-250, 1938.

(40) Tomkins, R. G. Organoleptic Tests in the Food Industry. Chem. and Indus. No. 11, 167-168, 1949.

(41) Weckel, K. G. Put Your Taste Buds to Work on Your Flavor Problems. Internat'l. Assoc. Milk Dealers Bull. 34, 136-140, 1941.

(42) Ziegler, E. C. and Schofield, E. H. Properties of Foods. Food Manufacture 20, 55, 1945.